약학박사학위논문

건강보험심사평가원 보험청구자료를 대상으로 한 약물상호작용 탐지 방법 적용 연구

Application of a drug-interaction detection method to the Korean National Health Insurance claims database

2014년 2월

서울대학교 대학원

약학과 예방·임상약학 전공

최 청 암

약학박사학위논문

건강보험심사평가원 보험청구자료를 대상으로 한 약물상호작용 탐지 방법 적용 연구

Application of a drug-interaction detection method to the Korean National Health Insurance claims database

2014년 2월

서울대학교 대학원

약학과 예방·임상약학 전공

최 청 암

건강보험심사평가원 보험청구자료를 대상으로 한 약물상호작용 탐지 방법 적용 연구

Application of a drug-interaction detection method to the Korean National Health Insurance claims database

지도교수 신 완 균

이 논문을 약학박사학위논문으로 제출함

2013년 11월

서울대학교 대학원

약학과 예방·임상약학전공

최 청 암

최청암의 박사학위논문을 인준함

2013년 12월

위 원 장 오 정 미 (인)

부 위 원 장 김 은 경 (인)

위 원 이 명 걸 (인)

위 원 이 병 구 (인)

위 원 신 완 균 (인)

ABSTRACT

Application of a drug-interaction detection method to the Korean National Health Insurance claims database

Chungam Choi
Department of Pharmacy
The Graduate School
Seoul National University

Drug interactions (DIs) constitute a serious problem and are considered to contribute to 6-30% of all adverse events (AEs). The use of existing data, including claims data, is expected to be helpful in detecting unknown DIs by complementing conventional spontaneous reporting systems. In the present study, an 'Ω shrinkage measure' was applied to the Korean National Health claims database to test the potential of the claims database as a DI surveillance resource. A well-known DI between non-steroidal anti inflammatory drugs (NSAIDs) and diuretics was analyzed using the model. International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) codes related to DIs were assigned to the AEs of the DIs: I50, I50.0, I50.1, I50.9, R60, R60.1, R60.9, and J81. An elevated occurrence of AEs versus the expected level was observed

using a two sided 95% lower credibility interval limit above zero, Ω_{025} = 0.245, which was the screening limit. The result was consistent with the actual DI between the two drugs. The finding indicates that the claims data have the potential to be used as a DI surveillance resource and that the Ω shrinkage measure may be a promising tool for detecting DIs in claims data.

Keywords: drug interaction, drug toxicity, data mining, health insurance, NSAIDs, diuretics

Student Number: 2008-21823

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

A. INTRODUCTION

Adverse events (AEs) constitute a serious problem and are a leading cause of drug-related death (Lazarou et al., 1998). In particular, because it is common for the chronically ill to be treated with polypharmacy, the unpredictability of drug use is ever increasing. One of the problems that can arise with multiple drug use is drug interactions (DIs) in which one drug influences the effects of another, primarily by increasing or decreasing its effects. Patients can suffer harm, both from excessive effects of a drug and from the neutralization of a medicine that is necessary for the patient. It is believed that 6-30% of AEs are due to DIs (Pirmohamed and Orme, 1998). However, detection of DIs remains challenging due to a lack of understanding of pharmacological mechanism(s) and practitioners' occasional misjudgments of DIs as simple AEs.

The use of post-marketing surveillance data has been emphasized in the early detection of AEs due to DIs. Among such post-marketing surveillance data, a major source of DI detection is AE case reports. AE case reports are mostly submitted spontaneously by healthcare professionals or patients. The submitted reports are generally collected and processed to be analyzed for detection of AEs. The most extensive post-marketing AE database is 'Vigibase', which is maintained by a collaboration of the Uppsala Monitoring Centre (the UMC) and World Health Organization (WHO) (Hugman, 2004). Vigibase is a computerized database containing AE case reports from member countries worldwide.

Several attempts have been made to use AE case reports from

Vigibase and other AE case-report databases for DI research. van Puijenbroek and colleagues investigated DIs between oral contraceptives and itraconazole by calculating the AE reporting odds ratios (van Puijenbroek et al., 1999) and examined the effects of pairs of NSAIDs with diuretics using a logistic regression model with AE case reports (van Puijenbroek et al., 2000). Another study was done to compare a multiplicative model and an additive model using known interactions of drugs (Thakrar et al., 2007). A group at UMC proposed a model with additive risk, namely, the 'Ω shrinkage measure,' which calculates an observed-to-expected ratio as a measure of disproportionality. This approach was adopted for AE case reports after its evaluation using known DIs (Noren et al., 2008; Qian et al., 2010).

Most efforts to detect AEs, including DIs, have focused on AE case reports. However, a movement has emerged to use other databases as sources for AE detection. In 2007, the U.S. Food and Drug Administration (FDA) initiated the New Sentinel Network, which uses multiple existing data sources, including health insurance claims data and electronic medical records (EMR) to complement the system that uses AE case reports (Platt et al., 2009). If the system is fully developed, both the early detection of signals and confirmation of signals from other sources are expected by virtue of the massive scale of the available data.

Korea has implemented an obligatory health insurance system managed by the Health Insurance Review and Assessment Service

(HIRA), which covers the Korean population. In the database of this system, enormous amounts of data on up to 50 million patients are collected nationally on a regular basis. Some researchers have tried to take advantage of this massive data set to detect AE signals (Choi et al., 2011; Choi et al., 2010; Kim et al., 2011). However, no attempt has been made to assess DIs using the database.

In this study, we examined the potential of using the HIRA as a DI surveillance source by applying an analysis method to data on a known DI case, which was processed from the HIRA. The method was developed by Noren et al. and was originally proposed for analyzing DIs in AE case reports. As a known interaction example case, DIs between NSAIDs and diuretics were analyzed. Administration of non-steroidal anti-inflammatory agents (NSAIDs) can inhibit the synthesis of prostaglandins in the kidney, which can cause sodium and water retention and hence diminish the effectiveness of diuretics (Brater et al., 1980; Clive and Stoff, 1984; Heerdink et al., 1998; Herchuelz et al., 1989; Schlondorff, 1993).

B. MATERIALS & METHODS

1. Data source

In Korea, all healthcare service claims are submitted to the HIRA by healthcare providers. Then, the HIRA determines the amount of reimbursement by reviewing the submitted healthcare service records. The assessed information is computerized and stored in the database, enabling analysis at a massive scale. Submitted research designs are reviewed by a committee, and only researchers of approved studies are allowed to access HIRA data within a restricted range. The data include information on national health insurance coverage: general information on services, medical treatments, diagnoses, and outpatient prescriptions. We used patient information, drug codes, diagnosis codes, and dates of prescribing in this study. Patient information covered age, gender, and an identification number, which was assigned arbitrarily. Drugs and diagnoses were encoded according to the Anatomical Therapeutic Chemical (ATC) classification system and the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10), respectively. We extracted 1 year of outpatient data (from 1 Jan 2008 to 31 Dec 2008) with additional diagnosis data for 7 months (from 1 Jun 2007 to 31 Dec 2008) from the HIRA database. Subjects were those aged over 17 years who visited a medical care institution during a 3-week period, between 1 and 21 June 2007, and who therefore had a record of receiving medical service. Codes in the claims data that were not in the list of codes or an age of over 130 years were regarded as errors. All the data of insurants containing errors were excluded.

2. Drug-interaction example case

We chose a well-known DI to evaluate the model to be applied to the HIRA data. In the interaction, ICD-10 codes that matched AEs derived from DIs were chosen. A physician checked and revised the list of ICD-10 codes, which were chosen by a pharmacist. The decreased effect of diuretics can be expressed as the occurrence of edema or heart failure, reflected by the following ICD-10 codes: I50, I50.0, I50.1, I50.9, R60, R60.1, R60.9, and J81 (Table 1). For the drugs, ATC codes starting with M01A and C03 were regarded as NSAIDs and diuretics, respectively.

Table 1. Chosen ICD-10 codes and their titles that match the ADRs for DIs between NSAIDs and diuretics

ICD-10	Title
I50	Heart failure
I50.0	Right ventricular failure (secondary to left heart failure)
I50.1	Left ventricular failure
I50.9	Heart failure, unspecified
J81	Pulmonary edema
R60	Edema (not elsewhere classified)
R60.0	Localized edema
R60.1	Generalized edema
R60.9	Fluid retention (not otherwise specified)

ADR, adverse drug reaction; DI, drug interaction; ICD-10, International Statistical Classification of Diseases and Related Health Problems, 10th Revision; NSAID, non-steroidal anti-inflammatory drug.

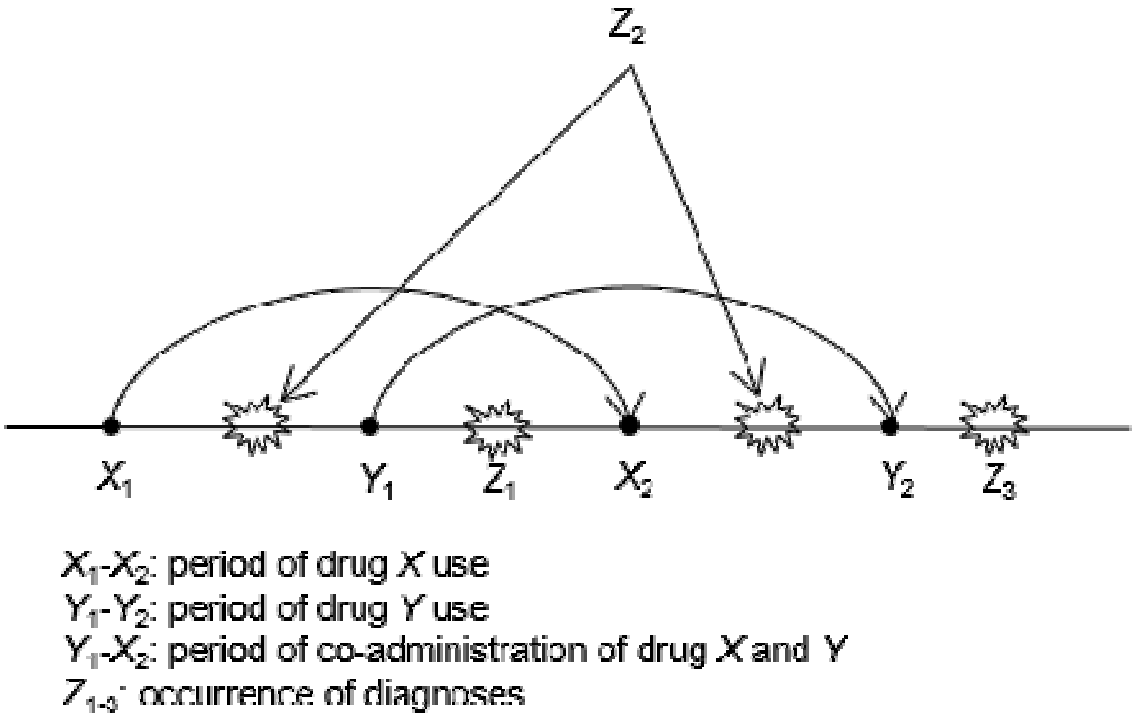

Figure 1. Description of the process used to generate drug-diagnosis pairs from HIRA data. HIRA, Health Insurance Review and Assessment Service

3. Data preprocessing

When patients visited clinics regularly to follow up on chronic diseases or to get a prescription, diagnoses recorded periodically could be mistaken for newly occurring diagnoses. To prevent this, it was important to eliminate records of repeated diagnoses. In our prior study using HIRA data, over 99% of the prescribed days of drugs were for no longer than 180 days. Considering the gap between actual visits and the expected day of visits as being up to 20 days, only diagnoses having no prior occurrence within 200 days were considered newly occurring diagnoses. Likewise, when a drug was prescribed again within 20 days after the administration period of the same drug ended, it was considered that the drug was used continuously, and the prescribing records were merged.

4. Model for DI detection

4.1. Generation of drug-diagnosis pairs that correspond to AEs

Unlike AE case reports, HIRA data consist of records of drugs and diagnoses. Drug-diagnosis pairs were used to calculate DI. The generation of drug-diagnosis pairs from HIRA data is shown in Figure 1. First, X and Y denote two drugs between which an interaction is known, and Z denotes all kinds of newly occurring diagnoses that are coded in ICD-10 codes. All of the events on the horizontal line were arranged in order of incident date. Z1 represents

a newly occurring diagnosis when the periods of drugs X and Y overlap, Z2 represents a newly occurring diagnosis when either X or Y is prescribed, and Z3 represents a newly occurring diagnosis when neither of the two drugs is prescribed. Each drug-diagnosis pair processed from HIRA data corresponded to variables marked with different signs: 0, 1, and ·, denoting negative, positive, and both, respectively. The first two signs gave us information as to whether drugs X or Y were used, and the last sign denoted whether the diagnostic codes of interest had occurred. We let n111 denote the count of drug-diagnosis pairs of both X and Y with diagnoses of interest. n11· denotes the count of both X and Y with all kinds of diagnoses, which includes n111. n101 denotes the count of drug-diagnosis pairs of only X with diagnoses of interest. Likewise, n10· denotes the count of all pairs of X without Y and so on.

4.2. Method for DI analysis

Noren et al. proposed a model of screening DIs in AE case reports. This method used a measure of disproportionality, the Ω shrinkage measure, which was based on an additive model. The calculation of Ω is

$$\Omega = \log_2 \frac{n_{111} + \alpha}{E_{111} + \alpha}$$

where E111 is the expected value of the incidence of disease suspected to be derived from DI, and α is a tuning parameter, which is set at 0.5 (Noren et al., 2008). From a Bayesian perspective, if both the prior and posterior distribution of μ are assumed to be gamma distributed, the exact limits of the credibility interval can be obtained by the following equation:

$$\int_0^{\mu_q} \frac{(E_{111}+\alpha)^{n_{111}+\alpha}}{\Gamma(n_{111}+\alpha)} u^{n_{111}+\alpha-1} e^{-(n_{111}+\alpha)u} du = q$$

A detailed description of this equation is provided by Noren et al. The logarithm of this equation for q = 0.025 is a two-sided 95% lower credibility interval limit, Ω_{025}, which acts as the threshold for screening DI when $\Omega_{025} > 0$ (Noren et al., 2008). This method was used in our study to process drug-diagnosis pairs from HIRA data. Bootstrapping was used to evaluate the accuracy and robustness of the model in HIRA data. Bootstrapping is acceptable for claims data, in that it allows all data to be included in model development and validation; other methods such as the split-sample approach lack reliability, and in case n111 the count is too small. The bootstrap process with 2,000 random draw repeats of drug-diagnosis pairs was used to calculate the Ω shrinkage measure of each sample. To validate the lower credibility limit, the 2.5th percentile value of the bootstrap data denoted by Ω_{boot} was examined to determine if it corresponded with Ω_{025} based on the threshold, 0.

C. RESULTS

As shown in Figure 2, 13,129,115 adult insurants visited a medical care institution in the period from 1-21 June, 2007. Of these, 1,565 insurants were excluded because of errors in their data. From the remaining 13,127,550 insurants, 575,131,274 prescription records for drugs and 288,304,638 newly occurring diagnoses were found. As shown in Table 2, a total of 862,497,889 pairs of drugs and diagnoses of interest were identified. The observed number of pairs with the diagnoses in Table 1 when NSAIDs and diuretics were co-administered (n111) was 91,592, and the expected number of such pairs (E111) was calculated to be 76,775.99. The logarithm of the observed-to-expected ratio (Ω) was 0.255, and the 95% lower credibility limit (Ω_{025}) was 0.245, which was over the screening limit value of the interaction. The 2.5th percentile of the Ω value from bootstrapping data (Ω_{boot}) was 0.247.

Table 3 shows all of the ICD-10 codes assigned DIs and the proportion of all drug-diagnosis pairs. Diagnoses related to fluid retention (R60, R60.0, R60.1, R60.9) were frequent and had higher incident proportions in the group containing both NSAIDs and diuretics compared to the other groups.

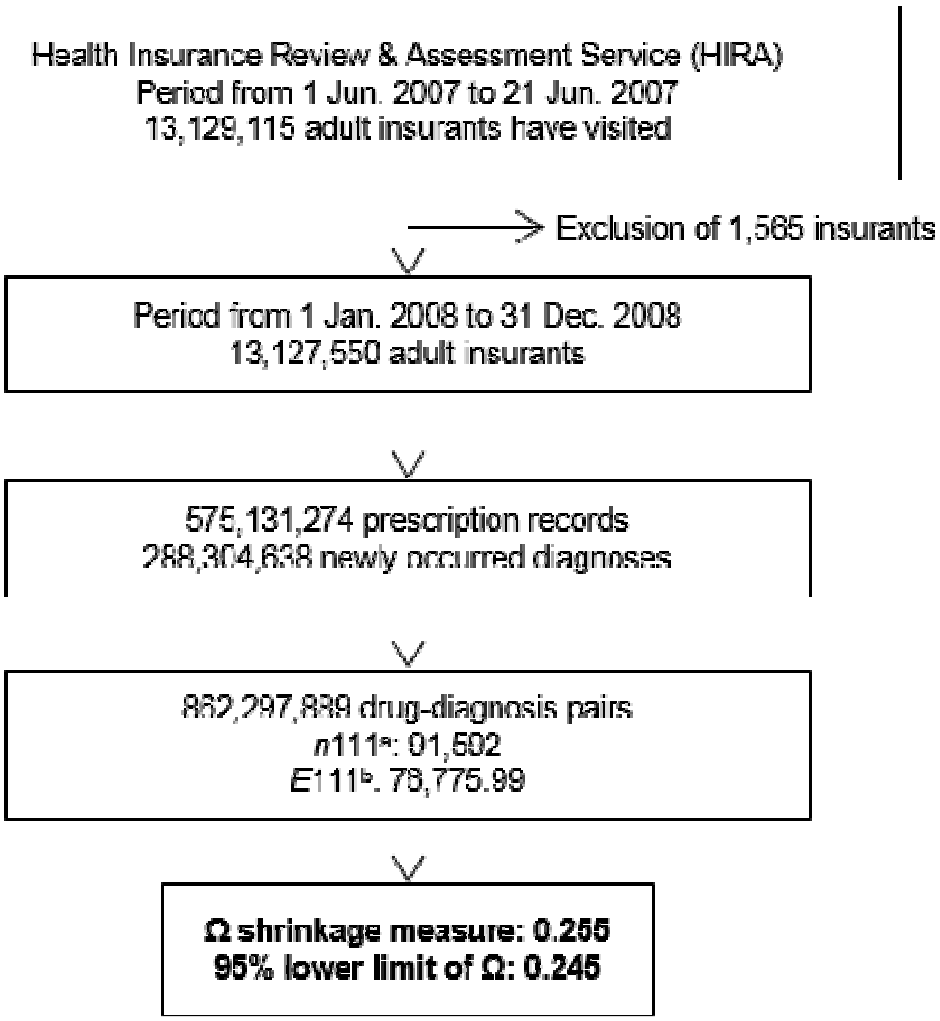

Figure 2. The process of calculating the Ω shrinkage measure from HIRA data. AE, adverse effect; HIRA, Health Insurance Review and Assessment Service. a Observed number of cases with the diagnoses shown in Table 1 when the records of NSAIDs and diuretics overlapped. b Expected number of cases with the diagnoses shown in Table 1 when the records of NSAIDs and diuretics overlapped.

Table 2. Drug interaction data for NSAIDs-diuretic interactions analyzed by Ω shrinkage measure using HIRA data

Variable	Number
$n_{00\cdot}$	789,523,883
n_{001}	3,937,895
$n_{10\cdot}$	47,613,053
n_{101}	207,760
$n_{01\cdot}$	17,436,553
n_{011}	168,935
$n_{11\cdot}$	7,924,400
n_{111}	91,592
E_{111}	76,775.99
Ω	0.255
Ω_{025}	0.245
Ω_{boot}	0.247

HIRA, Health Insurance Review and Assessment Service.

Table 3. Numbers of diagnoses by subgroups that were found in HIRA data that were assigned as DIs between NSAIDs and diuretics

Diagnosis	NSAIDs + Diuretics		NSAIDs only		Diuretics only	
	No. of occurring pairs	No./10,000 total pairs*	No. of occurring pairs	No./10,000 total pairs*	No. of occurring pairs	No./10,000 total pairs*
Heart failure	601	0.76	873	0.18	2,244	1.29
Right ventricular failure (secondary to left heart failure)	12,695	16.02	20,430	4.29	42,228	24.22
Left ventricular failure	467	0.59	952	0.20	1,864	1.07
Heart failure, unspecified	7,797	9.84	12,450	2.61	25,265	14.49
Pulmonary edema	1,572	1.98	3,081	0.65	5,960	3.42
Edema (not elsewhere classified)	932	1.18	2,118	0.44	1,800	1.03
Localized edema	16,367	20.65	45,895	9.64	20,072	11.51
Generalized edema	11,843	14.94	23,604	4.96	16,409	9.41
Fluid retention (not otherwise specified)	39,318	49.62	98,357	20.66	53,093	30.45
Total diagnoses	91,592	115.58	207,760	43.64	168,935	96.89

DI, drug interaction; NSAID, non-steroidal anti-inflammatory drug.

Total drug-diagnosis pairs of NSAID + Diuretics, NSAIDs only, and Diuretics only groups correspond to $n11\cdot$, $n10\cdot$, and $n01\cdot$, respectively.

D. DISCUSSION

The HIRA database is an extremely large database, consisting of the health information of millions of insured Koreans. With respect to DI research, the database has the possibility of discovering unknown DIs, which could be described as AEs in the package inserts, as well as unknown interactions of new drugs on the market. However, although some researchers have conducted drug-surveillance studies using HIRA data, there is presently neither any established method of analysis nor any sophisticated approach to processing these data. Furthermore, no previous research using HIRA data has focused on DI issues. This research is the first attempt to use HIRA data as a DI surveillance resource.

We applied a DI-analysis method to processed HIRA data to search for clues and to investigate the potential of this method for detecting DIs. The Ω shrinkage measure of the NSAID-diuretic interaction was 0.255, and its 95% lower credibility limit was above zero (Ω_{025} = 0.245) indicating that suspected DIs in the HIRA data occurred more frequently than expected when there was an NSAID-diuretic interaction. The bootstrapping result agreed with the model by showing that the 2.5th percentile of Ω was greater than 0 (Ω_{boot} = 0.247). The disproportion between n111 and E111 is indicative of a relationship between the overlap of the drugs and certain cardiovascular diagnoses. We were able to identify that the disproportionate occurrence corresponded to an actual interaction between the two drugs. This result showed that HIRA data can be a useful source for DI research.

Although the applied model was successful in analyzing a DI case, a few considerations must be given to applying this method to other cases, and fine tuning may be needed.

ICD-10 code selection is one of the most significant factors that should be considered, which can influence the application of the drug-interaction model. Inadequate selection can lead to a failure to capture interaction signals or, at worst, to an artifactual result. As shown in Table 3, not all of the ICD-10 codes reflected the actual interaction. While codes representing edema (except pulmonary edema) and fluid retention were in agreement with the actual interaction of the two drugs, other codes were not. Even when some codes were not in agreement, the overall tendency was towards a positive Ω_{025}, which results from the considerable weight of codes that agree. Likewise, there are risks that the selected ICD-10 codes failed to indicate AEs of interest. The discrepancy may be attributable to several factors, some of which cannot be completely disclosed when using claims data alone: differences in drug treatment duration, tendency of hospital visits among certain diseases, and possibilities of identifying certain diseases during routine check-ups. To obtain detailed insight on the code selection process, subsequent studies on other interactions and that take advantage of medical records as well as claims data are essential.

The basic assumptions of the model are another point of discussion. Several assumptions must be met to use this measure. Some

assumptions are in accordance with claims data: the assumption that n111 follows a Poisson distribution is reasonable in that n111 rarely occurs, and is independent to weakly dependent in either type of data. It can also be assumed that variation in generating drug-diagnosis pairs is negligible considering that the automated generation process has no preference for certain diagnoses or drugs.

However, other assumptions may lead to bias. First, products of probability have to be small enough so that they can be ignored in the calculation of the expected value (E111). While including a wide range of diagnoses is not likely to miss AEs, setting an excessive range can make the products of probabilities not negligible. The problem of estimability of background probability can arise according to selection of diagnoses. The average probability of a set of assigned diagnoses attributable to drugs of no interest should be small enough to estimate background probability for the diagnoses (Noren et al., 2008). While this can also be an obstacle when using AE case reports, there is a reason that more attention should be paid to claims data. Because drug-diagnosis pairs from HIRA data were defined simply by the time sequence of events without assessments by healthcare professionals, the pairs are more exposed to the risk of unknown relationships than to AE case reports, which can lead to violation of the basic assumption. Therefore, it is preferable to choose an AE for study that is considered to have a close, possibly exclusive, relationship with drugs of interest as well as appropriate ICD-10 codes to minimize the estimability issue.

There are advantages in using HIRA data as a DI surveillance source. The most distinctive merit is the enormous volume of data. Sufficient reports are needed in DI research to measure the observed-to-expected ratio. Korea has adopted a national insurance system that is mandatory, and information, including drugs and diagnoses, of all the insured are added to a computerized database regularly. Thus, as long as a drug is within insurance coverage, a large number of cases using that drug can be obtained. HIRA data also have an advantage in that there is no reporting bias concerning prejudice of reporters, which can affect the results of DI research (van der Heijden et al., 2002). Because suspected AEs are usually reported by healthcare professionals in practice, AEs that are well known or that are regarded as less serious have a chance of being ignored, whereas serious AEs or AEs exposed in the mass media may be reported more frequently. However, because processing of claims data is done by a computer, there is no likelihood of such reporting bias in studies using HIRA data.

This study has some limitations. Above all, the analysis of one case here does not confirm that the method is appropriate for analyzing all other DIs. To further ensure that the Ω shrinkage measure is appropriate as a DI analysis tool for HIRA data, other well-known interactions need to be analyzed. The study's limitations can be attributed in part to the nature of the database. HIRA data are based on the review and assessment of claims data. For this reason, the data may give less consideration to clinical information that does not

affect reimbursement, resulting in discrepancies between actual AEs and the selected set of ICD-10 codes. The wrong diagnostic codes may be recorded, which may affect the internal validity of the data. Healthcare professional routines are the main reason for wrong diagnoses (Kim, 2007). However, wrong diagnoses can arise from diagnoses made before adequate tests, which may remain as pseudo-diagnoses. Drug compliance can be a problem in a drug-surveillance study. There is no information about whether the patients took the drugs at the proper dosage throughout the whole period of the supposed administration.

DI surveillance research using the HIRA database is much more complex than surveillance of AE case reports. Unlike AE case reports, the data are analyzed using standard procedures. In the process, considerations arise at numerous points, ranging from mathematical issues to how healthcare professionals input their patient diagnoses. Nevertheless, massive health insurance claims databases are still a promising resource for drug-surveillance research.

E. CONCLUSION

HIRA data were processed for drug-diagnosis pairs using a DI-surveillance method. An interaction between NSAIDs and diuretics was identified using the Ω shrinkage measure, which reflected an actual interaction of the two drugs. The results showed that cases of a known DI could be detected by the Ω shrinkage measure, and HIRA data have the potential to be used as a resource for DI-surveillance research. However, several points need to be considered to apply this method to other DIs. Further studies applying the Ω shrinkage measure to other drugs are needed to ensure that the HIRA database is a reliable resource for detecting DIs.

F. REFERENCES

Brater, C., et al., 1980. Indomethacin and the response to bumetanide. Clin. Pharm. Ther. 27, 421-425.

Choi, N.-K., et al., 2011. Comparison and validation of data-mining indices for signal detection: using the Korean national health insurance claims database. Pharmacoepidemiology and Drug Safety. 20, 1278-1286.

Choi, N. K., et al., 2010. Signal detection of rosuvastatin compared to other statins: data-mining study using national health insurance claims database. Pharmacoepidemiology and Drug Safety. 19, 238-246.

Clive, D. M., Stoff, J. S., 1984. Renal syndromes associated with nonsteroidal antiinflammatory drugs. New England Journal of Medicine. 310, 563-572.

Heerdink, E. R., et al., 1998. NSAIDs associated with increased risk of congestive heart failure in elderly patients taking diuretics. Arch Intern Med. 158, 1108-1112.

Herchuelz, A., et al., 1989. Interaction between nonsteroidal anti-inflammatory drugs and loop diuretics: modulation by sodium balance. Journal of Pharmacology and Experimental Therapeutics. 248, 1175-1181.

Hugman, B., Viewpoint Part 2. the Uppsala Monitoring Centre, Uppsala, Sweden, 2004, pp. 14-25.

Kim, J., Using national health insurance data to vitalize the

evidence-based health care: the current status and tasks. The Korean Society for Preventive Medicine Winter Symposium, Seoul, Korea, 2007, pp. 1-28.

Kim, J., et al., 2011. Signal detection of methylphenidate by comparing a spontaneous reporting database with a claims database. Regulatory Toxicology and Pharmacology. 61, 154-160.

Lazarou, J., et al., 1998. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. Journal of the American Medical Association. 279, 1200-1205.

Noren, G. N., et al., 2008. A statistical methodology for drug-drug interaction surveillance. Statistics in Medicine. 27, 3057-3070.

Pirmohamed, D., Orme, M. L., 1998. Drug interactions of clinical importance. in: Davies, D., et al., (Eds.), Davies's Textbook of Adverse Drug Reactions. Chapman & Hall Medical, London, pp. 888-912.

Platt, R., et al., 2009. The new sentinel network - improving the evidence of medical-product safety. New England Journal of Medicine. 361, 645-647.

Qian, Y., et al., 2010. A computerized system for detecting signals due to drug-drug interactions in spontaneous reporting systems. British Journal of Clinical Pharmacology. 69, 67-73.

Schlondorff, D., 1993. Renal complications of nonsteroidal anti-inflammatory drugs. Kidney Int. 44, 643-653.

Thakrar, B. T., et al., 2007. Detecting signals of drug-drug interactions in a spontaneous reports database. British Journal of Clinical Pharmacology. 64, 489-495.

van der Heijden, P. G. M., et al., 2002. On the assessment of adverse drug reactions from spontaneous reporting systems: the influence of under-reporting on odds ratios. Statistics in Medicine. 21, 2027-2044.

van Puijenbroek, E. P., et al., 2000. Detecting drug-drug interactions using a database for spontaneous adverse drug reactions: an example with diuretics and non-steroidal anti-inflammatory drugs. European Journal of Clinical Pharmacology. 56, 733-738.

van Puijenbroek, E. P., et al., 1999. Signalling possible drug-drug interactions in a spontaneous reporting system: delay of withdrawal bleeding during concomitant use of oral contraceptives and itraconazole. British Journal of Clinical Pharmacology. 47, 689-693.

G. APPENDIX

Derivation of the expected value of drug-diagnosis pair ($E111$)

The derivation process by Norén et al. is shown below with modifications in terms of claims data in which AE cases were substituted by drug-diagnosis pairs. Observing prescription record, we considered situations of prescription records of two drugs denoted by X and Y: Either X or Y, overlap of X and Y, or neither of two drugs. Occurrence of diagnoses of interest were denoted by Z.

First, let a_0 denote background probability of Z, then the total probability p_{00} for a pair with neither X nor Y is the background probability:

$$p_{00} = \alpha_0$$

Eq. (A.1)

Let α_1 denote the probability of Z attributable to X, and let a_2 denote the probability attributable to Y. The total probability p_{10} for Z in pairs with X in the absence of Y is

$$p_{10} = 1-(1-a_0)(1-a_1) = a_0 + a_1 - a_0 \cdot a_1$$

Eq. (A.2)

Likewise, Total probability p_{01} is

$$p_{01} = 1-(1-\alpha_0)(1-\alpha_2)$$
Eq. (A.3)

The total probability of Z under overlapped prescription records of X and Y is

$$p_{11} = 1-(1-\alpha_0)(1-\alpha_1)(1-\alpha_2)$$
Eq. (A.4)

Given that both the background probability, α_0 and the attributable probability from X, a_1,can be assumed to be small for any diagnosis Z, their product, $a_0 \cdot a_1 \ll a_0, a_1$.Thus,the following approximation of Eq.(A.2) is valid:

$$p_{10} \approx \alpha_0 + \alpha_1$$
Eq. (A.5)

Similarly,

$$p_{01} \approx \alpha_0 + \alpha_2$$
Eq. (A.6)

$$p_{11} \approx \alpha_0 + \alpha_1 + \alpha_2$$
Eq. (A.7)

Let Z' denote the occurrence of at least one of the (potentially large) group of diagnoses excluding Z (and in its absence so that Z and Z' are mutually exclusive events). Let a'_0 denote the background probability of A'. If diagnoses with an attributable probability from either X or Y can be excluded from Z', the total probability of Z' will be a'_0 for all possible pairs of X and Y:

$$p'_{00} = a'_0$$
$$p'_{10} = a'_0$$
$$p'_{10} = a'_0$$
$$p'_{11} = a'_0$$

Eq. (A.8)

However, the identification of an appropriate set of unrelated diagnosis terms for a given pair of drugs requires expert clinical judgment, which cannot easily be automated for routine screening purposes. Common practice in pairwise disproportionality analysis of ADR surveillance data is therefore to include all ADRs other than Z in Z' for first pass screening purposes. Norén et al. proposed that the same approach be used for interaction screening since Eq. (A.8) will hold approximately unless X or Y considerably alters the overall risk of any suspected diagnosis in association with the prescription. Should this be the case, restriction of Z' to a more narrow set of diagnoses will resolve the problem.

Let

$$f_{00} = \frac{n_{001}}{n_{00\cdot}}$$

$$f_{10} = \frac{n_{101}}{n_{10\cdot}}$$

$$f_{01} = \frac{n_{011}}{n_{01\cdot}}$$

$$f_{11} = \frac{n_{111}}{n_{11\cdot}}$$

Eq. (A.9)

Denote the corresponding observed relative drug-diagnosis pair rate for Z.

Now, an estimator for the expected relative pair rate of Z under the overlapped use of X and Y (f_{11}), based on the relative drug-diagnosis pair rates of Z, can be constructed, given pair of at most one of X and Y (f_{00},f_{10}, and f_{01}).In order not to let potential interaction contaminate the estimation of the expected relative, it is exclusively based on f_{00},f_{10}, and f_{01}.The expected value for the background relative pair rate of Z in the absence of both X and Y is

$$E[f_{00}] = E[E[f_{00}|n_{00\cdot}]]$$

$$= E\left[\frac{\alpha_0}{\alpha_0 + \alpha'_0}\right]$$

$$= \frac{\alpha_0}{\alpha_0 + \alpha'_0}$$

Eq. (A.10)

Similarly,

$$E[f_{10}] = \frac{\alpha_0 + \alpha_1}{\alpha_0 + \alpha_1 + \alpha'_0}$$

Eq. (A.11)

$$E[f_{01}] = \frac{\alpha_0 + \alpha_2}{\alpha_0 + \alpha_2 + \alpha'_0}$$

Eq. (A.12)

$$E[f_{11}] = \frac{\alpha_0 + \alpha_1 + \alpha_2}{\alpha_0 + \alpha_1 + \alpha_2 + \alpha'_0}$$

Eq. (A.13)

After re-expression of Eq. (A.13) in terms of Eq. (A.10)-Eq. (A.12),

$$\begin{aligned} E[f_{11}] &= \frac{\alpha_0 + \alpha_1 + \alpha_2}{\alpha_0 + \alpha_1 + \alpha_2 + \alpha'_0} \\ &= 1 - \frac{\alpha'_0}{\alpha_0 + \alpha_1 + \alpha_2 + \alpha'_0} \\ &= 1 - \frac{1}{\frac{\alpha_0 + \alpha_1}{\alpha'_0} + \frac{\alpha_0 + \alpha_2}{\alpha'_0} - \frac{\alpha_0}{\alpha'_0} + 1} \\ &= 1 - \frac{1}{\frac{E[f_{10}]}{1 - E[f_{10}]} + \frac{E[f_{01}]}{1 - E[f_{01}]} - \frac{E[f_{00}]}{1 - E[f_{00}]} + 1} \end{aligned}$$

Eq. (A.14)

Thus, an estimator of $E[f_{11}]$, we may use

$$g_{11} = 1 - \frac{1}{\frac{f_{10}}{1-f_{10}} + \frac{f_{01}}{1-f_{01}} - \frac{f_{00}}{1-f_{00}} + 1}$$

However, in order to avoid possible misleading influence of negative a_1 or a_2 estimates, g_{11} is modified as follows:

$$g_{11} = 1 - \frac{1}{\frac{f_{10}}{1-f_{10}} + \frac{f_{01}}{1-f_{01}} - \frac{f_{00}}{1-f_{00}} + 1}$$

$$= 1 - \frac{1}{\max\left(\frac{f_{00}}{1-f_{00}}, \frac{f_{10}}{1-f_{10}}\right) + \max\left(\frac{f_{00}}{1-f_{00}}, \frac{f_{01}}{1-f_{01}}\right) - \frac{f_{00}}{1-f_{00}} + 1}$$

Eq. (A.15)

When $f_{10}<f_{00}$ (indicating no probability of Z attributable to X), this yields the most sensible estimator $g_{11}=\max(f_{00}, f_{01})$ and vice versa when $f_{01}<f_{00}$.

Finally, E_{111} can be obtained by multiplying g_{11} and $n_{11\cdot}$:

$$E_{111} = g_{11} \cdot n_{11\cdot}$$

Eq. (A.16)

H. ABSTRACT IN KOREAN

건강보험심사평가원 보험청구자료를 대상으로 한 약물상호작용 탐지 방법 적용 연구

약물상호작용은 심각한 문제로 여겨지고 있으며 모든 약물유해반응 중에서 6-30%를 차지한다는 연구가 있다. 따라서 약물상호작용 연구가 중요함에도 불구하고 단독 약물 유해사례 연구에 비해 약물간의 기전을 이해하는 등 고려할 점이 많아서 약물 상호작용 연구에 어려움이 있었다. 이러한 어려움을 해결하는 한 가지 수단으로 자발적보고자료를 이용하고자 하는 움직임이 나타났다. 자발적보고자료를 이용하는 연구방법은 세계보건기구(WHO)와 웁살라모니터링센터(the UMC)의 협력에 의해 특정 약물의 유해사례 시그널을 탐지 하는 데 주로 사용되었다. 그리고 상호작용 연구에서도 자발적보고자료를 이용한 모델에 대한 연구가 수행되었다. 이러한 자발적보고자료 외에도 보험청구자료를 포함한 기존의 자료들을 이용하는 것이 주목을 받으면서 일반적으로 활용하는 자발적보고자료를 이용한 연구를 보완할 수 있을 것으로 생각되었다. 본 연구에서는 자발적보고자료를 이용하여 연구가 수행된 바 있는 'Ω shrinkage measure'가 건강보험심사평가원 보험청구자료에 적용되어 보험청구자료가 약물상호작용 연구 자원으로 이용될 수 있는지에 대한 가능성을 확인해 보았다.

잘 알려진 약물상호작용인 비스테로이드성 소염진통제(NSAIDs)와 이뇨제의 상호작용을 이 모델을 이용하여 분석하였다. NSAID의 투여는 신장에서의 프로스타글란딘의 합성을 저해하여 수분과 염분의 저류를 유발하고 이는 이뇨제 효과의 감소를 유발시킨다는 것이 알려졌다. 보험청

구자료에서는 환자식별정보, 내원일, 외래처방일, 상병명, 약물코드, 약물투여기간 정보를 사용하였다. 약물상호작용과 연관된 질병·상해및사인통계분류규정 10판 (ICD-10)의 코드는 상호작용의 유해반응으로 지정되었다. 그 목록은 다음과 같다: I50, I50.0, I50.1, I50.9, R60, R60.1, R60.9, J81. 약물코드로는 세계보건기구 표준인 ATC코드를 이용하고 비스테로이드성 소염진통제와 이뇨제를 각각 M01A와 C03으로 시작되는 약물로 간주하였다. 보험청구자료에서 2007.6.1.-2007.6.21. 까지 요양기관에 내원한 기록이 있는 환자를 대상으로 그들의 2008.1.1.-2008.12.31. 동안의 자료를 이용하였다. 2007.6.1.-2007.12.31. 동안의 상병명 정보를 받아서 환자들의 기존 질환을 확인하는데 사용하였다. 외래환자의 약물정보 자료를 이용하였으며 내원 및 외래환자의 상병명 자료를 이용하였다. 같은 환자에서 20일 이내에 동일 약물을 재처방 받았을 경우 지속적인 사용으로 간주하였으며 200일 이내에 같은 환자에서 같은 상병명이 발생한 경우 기존 질환으로 간주하였다. NSAIDs나 이뇨제 중 한 약물의 단독 처방일부터 처방기간동안 기존 질환이 아닌 새로 발생한 상병명이 있을 경우 약물과 짝을 지었고 두 약물이 같이 처방된 기간 동안 새로 발생한 상병명이 있을 경우 두 약물과 동시에 짝을 지었다. X약물과 Y약물이 처방되었다고 가정하면, n111은 약물-상병명 짝에서 두 종류의 약물이 동시에 처방되고 지정한 상병명이 발생한 경우의 총 건수이다. n11·은 두 종류의 약물이 모두 처방된 발생한 모든 종류의 상병명의 총 건수이다. 이는 n111에서 발생한 약물-상병명 짝을 포함한다. n101은 X약물이 처방되고 Y약물이 처방되지 않았을 경우 지정한 상병명이 발생한 경우의 총 건수이다. 위와 같이 n10·은 X약만 처방되었을 경우 발생한 모든 약물-상병명 짝의 총 건수이다. 모든 경우를 고려한 짝의 개수를 센 다음 이를 이용하여 두 약물이 같이 사용되었을 경우의 기댓값이 산출되었

고 베이지안 통계에 의해 분석되었다. Gamma(0.5, 0.5)를 Ω의 사전분포로 놓고 사후분포를 추정하였으며 사후분포 신용구간의 양측 95% 하한치 값이 0보다 클 경우를 불균형으로 판단하였다. 타당성을 검증 하기 위해 bootstrap method를 이용하였다. 약물과 상병명의 짝을 대상으로 2000번의 무작위 추출을 하였으며 bootstrap 구간의 2.5 퍼센타일 값이 0을 초과하는지 확인하였다.

Ω의 95% 신용구간(credibility interval) 하한치가 스크리닝 한계인 0을 넘음으로써(Ω_{025} = 0.245) 유해사례가 기대치보다 더 많이 관측되는 것이 확인되었다. 상병명의 발생 빈도수를 환인해 본 결과 NSAIDs와 이뇨제가 같이 처방 되었을 경우 91,592 건의 유해사례 지정 상병명으로 10,000건당 총 115.58건이 산출되었고 NSAIDs만 단독으로 처방이 나왔을 경우엔 207,760건으로 10,000건당 43.64건, 이뇨제만 단독으로 처방이 나왔을 경우엔 168,935건으로 10,000건당 96.89건이 산출 된 것을 확인 할 수 있었다. 이는 주로 수분저류와 관련한 상병명에 기인한 것이었다.

본 연구에서 ICD-10 코드의 선택은 매우 중요한 과정이란 것을 확인하였다. 적절하지 않은 코드 선택은 상호작용을 탐지의 실패를 뿐만 아니라 위양성 시그널을 발생시킬수도 있다. 본 연구에서도 부종과 체내수분저류와 관련된 코드와 심부전과 관련된 코드 간에 결과 값에서 차이를 보이는 것을 확인 할 수 있었다. 이러한 원인은 몇 가지 요인에 의해서 유발될 가능성이 있는데 약물 치료 기간의 차이, 질병 간 요양기관 내원 패턴의 차이, 질병 간 주기적인 방문 및 검사 시 발견될 가능성의 차이 등이 있는 것으로 추정된다. 따라서 보험청구자료와 함께 의료기록을 같이 이용하는 추가적인 약물상호작용 연구를 수행하는 것이 필수적인 과제이다.

모델의 기본 가정에 대해서도 논의될 여지가 있다. 몇 가지 가정이 본 모델에 부합되어야 하는데 n111의 포아송 분포, 각 사건 간 독립성, 약물-진단 짝의 생성에 대한 분산에 대한 가정은 합리적이라 볼 수 있다. 하지만 확률간에 곱한 값이 무시할 수 없는 경우, 상병명이 알 수 없는 이유로 발생할 확률에 대한 가정에 대한 불일치는 모델에 비뚤림을 발생시킬 수 있다.

심평원자료를 이용한 상호작용 연구는 데이터의 양이 방대하고 다양한 약물과 질병에 대한 자료가 있어서 여러 가지 질병과 약물에 대한 약물감시 연구가 가능하다는 장점이 있다. 그리고 보고자나 약물에 대한 이슈에 따른 보고 비뚤림에서 자유롭다는 장점이 있다.

본 연구는 한 가지의 약물 상호작용에 대해서만 연구를 수행하였다. 따라서 보험청구자료에서 본 모델을 이용하기 위해서는 다른 잘 알려진 상호작용에 대해서도 추가연구가 필요하다. 그리고 보험청구자료의 자료 속성상 환자의 임상적인 정보에 대한 제한이 있기 때문에 잘못된 시그널이 발생할 수 있는 가능성이 있으며 의사의 진단코드 입력 부정확, 환자의 복약이행여부에 따른 연구의 한계점도 존재한다.

NSAIDs와 이뇨제를 대상으로 한 상호작용 모델 적용 연구에서 실제 두 약물간의 유해반응과 일치하는 결과를 나타내었다. 이 결과로 보험청구자료가 약물상호작용 연구의 자원으로 사용될 가능성이 있다는 것을 확인하였다. 한가지의 약물상호작용을 대상으로 연구하는 등 한계점은 있지만 Ω shrinkage measure가 보험청구자료에서 약물상호작용을 찾아내는데 유용한 도구로 사용 될 가능성이 있다는 사실을 확인하였다.

주요어: 약물 상호작용, 유해사례, 데이터마이닝, 건강보험, 비스테로이드성 소염진통제, 이뇨제

학 번: 2008-21823

I. CURRICULUM VITAE

Personal Profile

Name: Chungam Choi

Date of Birth: Apr. 20 1985

Office Tel.: +82-2-740-8556

Fax.: +82-2-766-8556

E-mail: aamssa04@snu.ac.kr

Educational Background

2008.3.1 - 2014.2.26. Ph.D., Majored in Pharmacy, College of Pharmacy, Seoul National University, Seoul, Korea

2004.3.1 - 2008.2.26 B.S., Majored in Pharmacy, College of Pharmacy, Seoul National University, Seoul, Korea

2001.3.1 - 2004.02.24 Diploma, Pohang high school, Pohang, Korea

Licence

Pharmacist #60941, 2008.3.7

Training

2010.8.24 - 2010.8.26 2010 NONMEM for PK/PD Modeling (Workshop), Chungnam National University, Daejeon, Korea.

2012.2.2 - 2012.2.5 The 5th NONMEM Workshop, Department of Clinical Pharmacology and Therapeutics, Asan Medical Center, Seoul, Korea.

2013.5.3 - 2013.5.4 Cochrane Review Workshop held by The Korean Branch of the Australasian Cochrane Center, Korea university, Seoul, Korea.

2013.7.11 Research Ethics Workshop, Seoul National University, Seoul, Korea.

Professional Experience

2008.3.1 - 2012.3.31 Researcher, Evaluation of the safety of functional food, Korea Ministry of Health and Welfare.

2008.3.1 - 2008.4.30 Researcher, Construction of the drug information web site for general public (http://medication.kfda.go.kr), Korea Food and Drug Administration.

2008.9.26 - 2009.9.30 Research manager, Bioequivalence test of Acetaminophen extended release tablet, Hana

Pharmaceutical Co., Ltd.

2009.6.24 - 2011.11.30 Researcher, The development of signal detection research using administrative data for public insurance, Korea Food and Drug Administration.

2011.8.9 - 2012.2.28 Researcher, Research on safe administration of self-administered biomedicine, Korea Food and Drug Administration.

2012.2.15 - 2012.11.30 Researcher, Development of Database for Advancement of Utilizing the Provision of Patient Education for Drugs (II), Korea Food and Drug Administration.

2013.2.1 - 2013.11.30 Researcher, Research on safe administration of self-administered biomedicine (II), Korea Food and Drug Administration.

Patent

Shin WG, Choi CA, #1020110146824, Signal detection system and method for adverse drug reaction using insurance claims data, Korea, 2011.12.30.

Publications

Choi CA, Chang MJ, Choi HD, Chung WY, Shin WG, Application of a drug-interaction detection method to the Korean National Health Insurance claims database, Regul. Toxicol. Pharmacol., 2013, 67:294-298.

Choi HD, Lee HJ, Lee SH, Kim SH, **Choi CA**, Chang MJ, Shin WG, Pharmacokinetics and correlation analysis of cilostazol in healthy Korean subjects, Int J Clin Pharmacol Ther, 2011, 50(5):345-348.

Kim JH, **Choi CA**, Oh JM, Son SH, Shin WG, Use of Information Component (IC) and Relative Risk (RR) for Signal Detection of Drug Interactions of Clopidogrel : Data-mining Study Using Health Insurance Review & Assessment Service (HIRA) Claims Database, Kor. J. Clin. Pharm., 2011, 21(2):90-99.

Presentations

Oral

Choi CA, Shin WG, Adverse drug reaction (ADR) Signal detection research using administrative data for public insurance, 2012 International Congress of Korean Federation of Pharmaceutical Societies, Jeju, Korea, 19-21 Apr 2012.

Poster

Choi CA, Kim JH, Chang MG, Choi HD, Shin WG, Singnal detection of drug interaction of clopidogrel in Korean national health insurance database, 11th Asian Conference on Clinical Pharmacy, Pasay City, Philippines, 24-27 Jun 2011.

Choi CA, Chang MJ, Choi HD, Shin WG, Adverse drug reaction (ADR) signal detection research using administrative data for public insurance, 12th Asian Conference on Clinical Pharmacy, Hong Kong, 7-9 Jul 2012.

Chang MG, Choi HD, **Choi CA**, Lee JH, and Shin WG, Therapeutic drug monitoring (TDM) of the antituberculosis drugs in patients with diabetes mellitus (DM), 2009 American Society of Health-System Pharmacists Midyear Clinical Meeting, 6-11 Dec 2009.

Eum SN, **Choi CA**, Kim HA, Shin WG, Signal detection of drug interaction of Tamiflu in Korean national health insurance review database, 10th Congress of the European Association for Clinical Pharmacology and Therapeutics, Budapest, Hungry, 26-29 Jun 2011.

Kim JH, **Choi CA**, Shin WG, Signal Detection of Adverse Drug Reaction of Clopidogrel Using Administrative Data for korean Public Insurance, International Conference Pharmacoepidemiology (ICPE), Brighton, UK, 19-22 Aug 2010.

Lee SH, **Choi CA**, Shin WG, Signal Detection Research of Oseltamivir Using Administrative Data for Public Insurance,

International Conference Pharmacoepidemiology (ICPE), Brighton, UK, 19-22 Aug 2010.

Kim SO, **Choi CA**, Kim JH, Chang MJ, Choi HD, Shin WG, Evaluation of applying data-mining indicators to national health insurance database by detecting an adverse effect(angle-closure glaucoma) of topiramate, 11th Asian Conference on Clinical Pharmacy, Makati City, Philippines, 24-27 Jun 2011.